



PENGENALAN

Semakan julat data (range check) merupakan proses penting dalam semakan kualiti data Banci Ekonomi yang dilaksanakan oleh Jabatan Perangkaan Malaysia (DOSM), untuk memastikan kualiti, ketepatan, dan kebolehpercayaan maklumat serta konsisten dengan reality operasi perniagaan sebenar.



APAKAH SEMAKAN JULAT DATA?

Semakan julat data ialah kaedah kawalan kualiti untuk memastikan nilai pemboleh ubah berada dalam had minimum dan maksimum yang munasabah atau ditetapkan. Nilai di luar julat biasanya menandakan:

- ralat kemasukan data,
- ralat pengkodan,
- salah unit pengukuran,
- atau kes terpencil (outlier) yang perlu disahkan.



PEGAWAI DOSM – SEMAKAN JULAT

JENIS JULAT DATA

a. Julat logik (logical range)

- Berdasarkan logik atau sifat pemboleh ubah
- Jumlah jam operasi: 0 - 24
- Umur perniagaan (tahun operasi): ≥ 0
- Bilangan cawangan/premis: ≥ 1 (jika syarikat aktif)

b. Julat ditakrif (defined range)

- Berdasarkan takrif pemboleh ubah/metadata
- Kod status: {1, 2, 3}
- Jenis pemilihan: 1 – 4

c. Julat empirikal (statistik)

- Berdasarkan pemerhatian data
- Mean ± 3 sisihan piawai
- Julat antara kuartil (IQR) - digunakan untuk mengesan outlier, bukan terus membuang data

TUJUAN SEMAKAN JULAT

- Mengesan nilai tidak munasabah/mustahil
- Meningkatkan ketepatan dan kebolehpercayaan data
- Mengelakkan kesilapan analisis dan pelaporan
- Menyeragamkan kualiti data merentas sumber atau siri masa

CONTOH SEMAKAN JULAT

Contoh 1: Bilangan pekerja

- Julat dibenarkan: ≥ 1
- Rekod bermasalah: -3, 0

Tindakan:

- Semak borang soal selidik
- Sahkan dengan sumber asal
- Betulkan atau tandakan sebagai missing

Contoh 2: Nilai jualan (RM)

- Julat munasabah: 0 – 100,000
- Nilai dikesan: 1,000,000

Kemungkinan isu:

- Salah unit (tahunan vs bulanan)
- Tersalah letak titik perpuluhan





PELAKSANAAN SEMAKAN JULAT (CONTOH TEKNIKAL)

Dilaksanakan menggunakan **peraturan yang ditetapkan**:

Julat Logik

- Mustahil dari segi realiti
- Contoh: Jumlah jam bekerja = -3, 25

Julat Ditakrif

- Tidak ikut metadata
- Contoh: Kod status = 7 (tidak sah)

Julat Empirikal

- Nilai luar biasa (*outlier*)
- Berdasarkan corak data sebenar

a. Dalam Excel

Gunakan **Data Validation**:

- Allow: Whole number/Decimal
- Minimum: nilai min.
- Maximum: nilai max.

Atau guna formula semakan:

=IF(ISNUMBER(MATCH(H2,{1,2,3,4},0)),"OK","ERROR")

b. Dalam SQL

```
SELECT *,
CASE
    WHEN jenispemilikan IS NULL THEN 'Missing'
    WHEN jenispemilikan NOT IN (1,2,3,4) THEN 'Error'
    ELSE 'OK'
END AS status
FROM data_responden;
```

c. Dalam Python

```
df['status'] = df['jenispemilikan'].isin([1,2,3,4])
df['status'] = df['status'].map({True: 'OK', False: 'Error'})
```

Discard (Buang)

Data **tidak boleh dipercayai** atau **terlalu rosak**, jadi ia **dibuang** dan tidak digunakan langsung.

Contoh:

- Nilai negatif untuk kuantiti jualan (tidak logik)
- Data terlalu ekstrem (*outlier*) tanpa justifikasi

→ Data ini **discard**

TINDAKAN SUSULAN SELEPAS SEMAKAN JULAT

OK

→ Teruskan ke analisis & tabulasi

ERROR

- Rujuk borang/responden
- Betulkan atau tandakan sebagai *missing*
- Rekod dalam log pembetulan

WARNING

- Disemak oleh pegawai *subjek-matter*
- Jangan *auto-edit*
- Kekalkan jika disahkan **benar**

Approve (Luluskan)

Data dianggap **betul**, **lengkap**, dan **boleh digunakan** tanpa perubahan

Contoh:

Nilai jualan = RM25,000 dan konsisten dengan rekod lain

→ Data ini **approve**

Impute (Isi/Anggar Nilai)

Data mempunyai **kekosongan** atau **ralat kecil**, jadi kita **isi** atau **anggar** nilai berdasarkan kaedah tertentu.

Kaedah biasa:

- Purata (mean)
- Median
- Nilai bulan sebelumnya
- Trend data

Contoh:

Data Februari tiada nilai → ambil purata Januari & Mac

→ Data ini **impute**

Ringkasnya: **Approve** → guna terus | **Impute** → isi/anggar nilai | **Discard** → buang data





FLOW SEMAKAN JULAT



AMALAN TERBAIK (BEST PRACTICES)

- ❑ Tetapkan julat dalam **metadata/data dictionary** berdasarkan data historikal (banci sebelumnya) dan data pentadbiran.
- ❑ Bezakan antara **ralat** dan **outlier sebenar**
- ❑ Laksanakan semakan julat:
 - semasa kemasukan data
 - sebelum analisis
 - sebelum diterbitkan
- ❑ Simpan log pembetulan data (audit trail)

